(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 3 013 013 A1

(12) EUROPEAN PATENT APPLICATION

(43) Date of publication:
27.04.2016 Bulletin 2016/17

(21) Application number: **14290318.6**

(22) Date of filing: **21.10.2014**

(51) Int Cl.:
H04L 29/06 *(2006.01)* ***H04N 21/24*** *(2011.01)*
H04N 21/439 *(2011.01)* ***H04N 21/44*** *(2011.01)*
H04L 12/861 *(2013.01)*

(84) Designated Contracting States:
AL AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO PL PT RO RS SE SI SK SM TR
Designated Extension States:
BA ME

(71) Applicant: **ALCATEL LUCENT**
92100 Boulogne-Billancourt (FR)

(72) Inventors:
- **Tang, M Siyu**
  2018 Antwerp (BE)
- **Rondao Alface, Patrice**
  2018 Antwerp (BE)
- **Vermoesen, Luc**
  2018 Antwerp (BE)

(74) Representative: **Plas, Axel Ivo Michel**
IP HILLS NV
Hubert Frère-Orbanlaan 329
9000 Gent (BE)

(54) Networking device and method for buffer management in video streaming over a network

(57) According to a first aspect, the invention relates to a networking device (100) for buffer management in video streaming over a network. The networking device (100) comprises a network interface (250) for receiving one or more video streams from a network (200). The one or more video streams comprise one or more low priority video streams containing low priority content and one or more high priority video streams containing high priority content. Further, the networking device (100) comprises one or more low priority buffers (502) for buffering packets of the respectively one or more low priority video streams and one or more high priority buffers (402) for buffering packets of the respectively one or more high priority video streams. The networking device (100) comprises a buffer synchronization agent (600) configured to monitor buffer delays of the buffers and configured to determine a common delay b* for the buffers (402, 502), wherein b* is smaller than the monitored low priority buffer delay if a monitored low priority buffer delay exceeds the monitored high priority buffer delays. The buffer synchronization agent (600) is configured to enforce the determined common delay b* to the buffers (402, 502). The networking device (100) further comprises a composer configured to combine the packets of the video buffers (402, 502) into frames of a combined video stream.

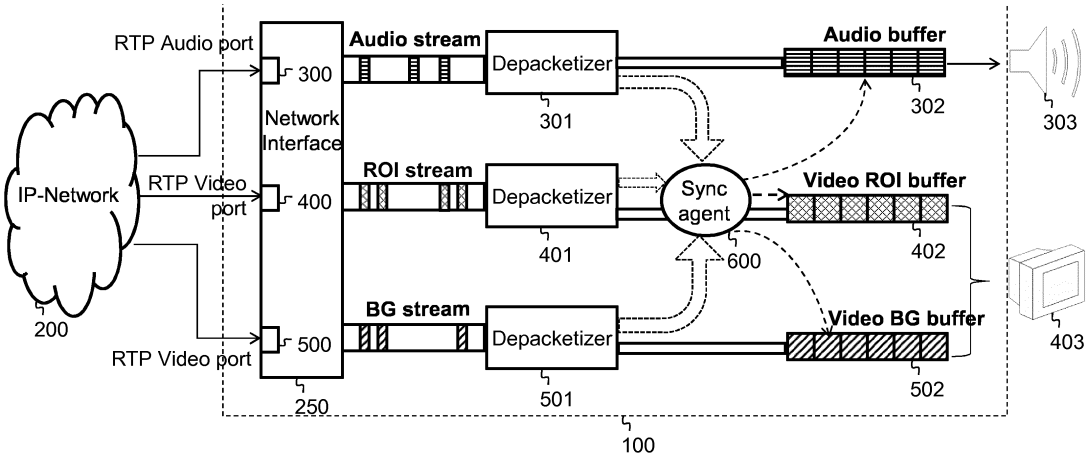


Fig. 1

EP 3 013 013 A1

Description

Technical Field

[0001] The disclosure relates to the field of network devices for video streaming over a network between a sender and a receiver, more particularly interactive and video streaming implementing inter-stream synchronization schemes and buffer management.

Background

[0002] Interactive and/or real-time communication is expected to be an important service in the next generation of multimedia delivery over a network (e.g. the Internet). A fundamental characteristic of today's real-time video conferencing traffic (e.g. Skype, Facetime, WebRTC, webcasting) is to share the same queue with a variety of Internet traffic, e.g. Web requests, bulk data transfer, adaptive streaming over HTTP, or other types of real-time flows.

[0003] Typically audio and video streams that are sent over the network, are buffered at the receiver. The information for these buffers need inter-stream synchronization or lip synchronization. Real-time audio/video traffic may be delivered via different networking paths and experience heterogeneous end-to-end (E2E) network delay or jitter when arriving at the receiver. Hence, audio/video streams need to be synchronized for play-out without noticeable audio/video skew. E.g. the minimum skew detectable by a human (between audio and video streams) is $\pm$80ms.

[0004] Different solutions are proposed to solve the lip-synchronization problem.

[0005] A first solution is lip-synchronization with the assistance of the RTCP Sender Report (RTCP SR) in "Real-time Transport Protocol" (RTP) communication over the network. The RTCP SR allows the receiver to map the RTP timestamp of each stream into a common reference timebase, called the Network Time Protocol (NTP) time, which is used for audio/video playback. Often, the usage of RTP timing information is not sufficient to provide inter-stream synchronization because the difference between the E2E delay of audio and video may vary substantially.

[0006] In a second solution, separate audio and video buffers are deployed for packet de-jittering. To provide inter-stream synchronization, either the audio or the video stream is delayed at the receiver via adaptive buffer management. An example is the master-slave synchronization. Existing buffer management schemes for inter-stream synchronization typically maintain an audio jitter buffer and a video play-out jitter buffer at the receiver. In a master-slave implementation, the audio stream is considered as the master. This means the video will be delayed based on the received audio stream in order to achieve lip synchronization. Notice that a large jitter buffer impairs the interactivity of the audio/video service, while a small buffer size may degrade the audio/video quality at the receiver. The play-out time of the video stream, serving as the slave, is passively adjusted to achieve inter-stream synchronization. A synchronization agent is placed between the audio and video buffer so as to exchange and negotiate the delay accordingly. By doing so, the audio and video buffers become time synchronized.

[0007] A third solution is the dynamic master-slave synchronization (e.g. Google's WebRTC), wherein both audio and video can serve as the master stream depending on the current buffer size of the audio and video play-out jitter buffers. In a first step, the current audio b_a and video b_v buffer size (in ms) is checked at the receiver. Then it is calculated how much later or earlier the audio stream is compared to the video stream, i.e., whether $b_a > b_v$ or $b_a < b_v$.

[0008] If the minimum video buffer delay is longer than the current audio buffer delay $b_a < b_v$, it further checks if extra delay has been added to the video engine in previous round-trip-times (RTT's). If so, the extra video buffer delay is reduced before adjusting the audio buffer delay. If there is no extra video delay to remove, the audio buffer delay is increased in accordance to the video buffer delay for inter-stream synchronization.

[0009] If the minimum video buffer delay is lower than the current audio buffer delay $b_a > b_v$, possible extra audio buffer delay is decreased or additional video buffer delay is added for synchronization. If there is extra delay in the audio buffer, the audio buffer delay is decreased, if there is no extra delay in the audio buffer, the video buffer delay is increased.

Summary

[0010] A problem with existing solutions is a need for better buffer management in real-time video stream receivers in view of interactivity and quality of a network connection. Although selecting the largest buffer delay ensures inter-stream synchronization, this may compromise the interactivity of the real-time service. Unnecessary buffer delay exists regardless of the importance of the content to be played-out.

[0011] This objective is achieved, in a first aspect, by a networking device for buffer management in video streaming over a network. The networking device comprises a network interface for receiving one or more video streams from a network. The one or more video streams comprise one or more low priority video streams containing low priority content and one or more high priority video streams containing high priority content. Further, the networking device comprises one or more low priority buffers for buffering packets of the respectively one or more low priority video streams and one or more high priority buffers for buffering packets of the respectively one or more high priority video streams. The networking device comprises a buffer synchronization agent configured to monitor buffer delays of the buffers.

[0012] In order to be able to drop packets with low-

priority content for interactivity optimization, the buffer synchronization agent is configured to determine a common delay b* for the buffers, wherein b* is smaller than the monitored low priority buffer delay if the monitored low priority buffer delay exceeds the other monitored buffer delays. The buffer synchronization agent is configured to enforce the determined common delay b* to the buffers. The networking device further comprises a composer configured to combine the packets of the video buffers into frames of a combined video stream.

[0013] The buffer synchronization agent monitors the buffer delay for the low and high priority video stream and coordinates these buffers. Instead of waiting for the arrival of a late packet with low priority content, e.g. packets from the low priority video stream, the agent decides to reduce the common buffer delay b* so as to improve the interactivity and drop these low priority packets or data. This will be executed if the monitored low priority buffer delay exceeds the other monitored buffer delays.

[0014] Hence, the common delay enforced on either of the jitter buffers makes a trade-off between the interactivity and quality degradation (cfr. distortion caused by the lost or late packets from the low priority video stream).

[0015] Diminishing packets of a low priority video stream is advantageous as this could be in favour of the quality of high priority video streams and in view of interactivity. For instance, a late video packet may correspond to less important information such as a static background. An extra delay will then not be added to the common buffer delay. Optionally, the distortion caused by the late packet can be reconstructed with error concealment. It is advantageous as video jitter buffers are adjusted adaptively based on the content (cfr. low or high priority) to be played out.

[0016] The term "networking device" in underlying disclosure should be understood as a computing device or computer which is provided with a network interface for communicating with one or more other devices via a network. Possible implementations of the networking device could comprise a computer, server, a media player, a video conferencing device, a portable device, etc.

[0017] The term "video stream" in underlying disclosure should be understood as video data, frames or packets which are continuously provided to the networking device via a network. Video data is played out at the receiver side, while video information is presented by the network to the networking device. Different video streams can be sent separately to the networking device via different network ports of the networking device, or can be aggregated and delivered through a common network port of the networking device.

[0018] One example of this video information could be captured video data which is provided in real-time to the networking device, e.g. live video streaming, interactive conferences, webcam streaming, etc. Another example could be the streaming of a recorded and stored movie. A video stream can relate to asynchronous video conversations, e.g. webinars, meeting recordings, online workshops, etc. Herein, video data is recorded and stored during a certain period, which video data can be viewed later and therefore be streamed over the network at a later instance.

[0019] The term "receiving a video stream over a network" in underlying disclosure should be understood as receiving packets or frames of a video stream provided or distributed via a network. Typically the packets are transported according to a standardized real-time media delivering protocol, such as e.g. "Real-time Transport Protocol" (RTP), "Real Data Transport" (RDT), "Real Time Streaming Protocol" (RTSP), H.323, "Session Initiation Protocol" (SIP), "MPEG Media Transport " (MMT), etc. The term "RTP port" in underlying disclosure should be understood as a port of a network interface suitable for communication via the RTP protocol or any other transport protocol as listed before.

[0020] The term "network path" in underlying disclosure should be understood as a temporally link or connection between a sender and the networking device. The network path is subject to a network delay, packet loss, etc.

[0021] The term "quality of a video stream" in underlying disclosure should be understood as the quality of frames or packets of the video stream, whereby the quality can be dependent on QoS (Quality of Service), an encoding level, spatial resolution, bitrate, PSNR (Peak Signal-to-Noise Ratio), etc.

[0022] The term "composer" in underlying disclosure should be understood as a device enabled to combine two or more video streams into a video stream or so-called combined video stream. A video player subsequently plays the combined video stream and/or an audio player synchronously plays the received audio stream accompanying the video stream. The received media (video and/or audio) can be played immediately without downloading the whole or chunks of video and/or audio stream. Video data, frames or packets of a low or high priority video stream comprise partial information of the combined video stream. In a preferred embodiment, video stream content corresponds to video data of a certain priority region.

[0023] The term "priority content" in underlying disclosure should be understood as a region, area, collection of areas or other content in a video frame with a certain associated priority. A high priority region could e.g. correspond to a talking person, while a low priority region could correspond to a static background. For instance, low and/or high priority content can be determined by video analysis based on background subtraction, motion estimation, temporal activity and/or saliency estimation. In another example, low and/or high priority content can be determined by the receiver or networking device.

[0024] The term "buffer", "playout buffer" or "(de)jitter buffer" in underlying disclosure should be understood as a buffer, which is used to counter jitter introduced by queuing in packet switch networks. Hence, this buffer allows for a continuous play-out of audio or video streams

transmitted over the network.

[0025] The term "monitored buffer delay" in underlying disclosure should be understood as the buffer delay or 'buffer size' as monitored or observed at a certain moment. This delay is determined by the buffer synchronization agent and is typically expressed in ms. For instance these values can be obtained by calling software programmed functions.

[0026] The term "common delay" in underlying disclosure should be understood as a delay which is enforced to and thereby common for the one or more low priority buffers, the one or more high priority buffers and audio buffer of the networking device.

[0027] The term "error concealment" in underlying disclosure should be understood as a technique whereby an error (e.g. packet loss) in a video stream is replaced by synthetic content, in an attempt to produce a more pleasant experience for the viewer. An example of error concealment could be an interpolation based on other content, frames or packets in the corresponding video stream.

[0028] According to an embodiment, the networking device further comprises an audio buffer for buffering packets of an audio stream, received from the network by the network interface, wherein the buffer synchronization agent is configured to monitor the buffer delay of the audio buffer, and wherein the buffer synchronization agent is configured to enforce the determined common delay b* to the audio buffer.

[0029] The audio stream or signal comprises relevant information and is important for the continuity of e.g. real-time streaming. The buffer synchronization agent monitors the buffer delays of the audio stream, low and high priority video streams and coordinates the jitter buffers in a way that packets with low-priority content are dropped for interactivity optimization. This means that, upon receiving the audio packets to be played out, the synchronization agent checks the video data that should be synchronized with the audio packets. Instead of waiting for the arrival of a late packet with low priority content, e.g. packets comprising static background content, the agent determines to reduce the common buffer delay so as to improve the interactivity.

[0030] According to an embodiment, the determined common delay b* also equals or exceeds the monitored buffer delay of the audio buffer, equals or exceeds the monitored buffer delays of the one or more high priority buffers and, equals or exceeds a minimum value b_{min}. Preferably, the minimum jitter buffer delay b_{min} is dependent on the buffering status or buffer delay of the audio buffer.

[0031] The play-out time of the audio stream is determined by keeping the intra-stream synchronization errors (e.g. packet de-jittering and audio temporal relationship reconstruction) as low as possible.

[0032] Therefore assigning higher priority to the audio stream than for the video streams, is advantageous as the audio stream or signal comprises relevant information and is important for the continuity of the conference.

[0033] The minimum jitter buffer delay b_{min} allows for a smooth playout of the audio stream.

[0034] According to an embodiment, the common delay b* maximally is determined to a value b_{max}, whereby b_{max} preferably is situated between 20ms and 150ms.

[0035] As the common buffer delay will not exceed b_{max}, impairment of interactivity is limited in a trade off with quality of the streams.

[0036] According to an embodiment, if a monitored low priority buffer delay exceeds the monitored high priority buffer delays, the buffer synchronization agent is adapted to determine the common delay b* sufficiently low while not exceeding a level of distortion do for the buffered low priority video stream and/or for the combined stream.

[0037] This way, a user or system could set a personalized level of distortion for a video streaming session, e.g. via a user interface. This allows for a more adaptive implementation for the trade-off between interactivity and video quality.

[0038] According to an embodiment, the networking device is configured to apply error concealment to the low priority content or frames of the combined video stream, if the determined common delay b* is smaller than a monitored low priority buffer delay and wherein the monitored low priority buffer delay exceeds the monitored high priority buffer delays.

[0039] As the common buffer delay is decreased at the expense of the monitored buffer delay of the low priority buffer, interactivity increases, but video quality could deteriorate. Error concealment is advantageous as it compensates for the decreased video quality.

[0040] According to an embodiment, the networking device is configured to provide the combined video stream and/or audio stream to a device suitable for playing the combined video stream and/or audio stream.

[0041] According to an embodiment, the network interface is configured to receive one or more streams of the one or more high priority video streams over the network via one or more ports of the network interface, separate from a port for receiving one or more streams of said one or more low priority video streams. Possibly the network interface is configured to receive the one or more streams of the one or more high priority video streams via one or more corresponding network paths with a lower congestion level than the network paths corresponding to the one or more low priority video streams.

[0042] This differentiated quality attribution allows to provide better quality to more prioritized regions and therefore high priority streams, in response of lower network congestion. Hereto, e.g. "differentiated services" (DiffServ) could provide differentiated QoS to the network paths. A "new payload type" of the transport protocol can be issued to identify both types of video streams.

[0043] According to an embodiment, the network interface is configured to receive the audio stream over the network via a port of the network interface, separate from one or more ports for receiving the video streams. Pos-

sibly the network interface is configured to receive the audio stream via a corresponding network path with a lower congestion level than the network paths corresponding to the video streams.

[0044] This differentiated quality attribution allows to provide better quality to the audio stream, which is important for the continuity of the video streaming or conference.

[0045] According to an embodiment, the networking device is configured to receive a high priority video stream with a higher quality than the one or more low priority video streams, in case of degraded network congestion. The quality can for instance correspond to an encoding level. Possibly the quality is inversely proportional to a congestion level of a network path of the network for the stream.

[0046] Indeed, encoding allows compression of video data and therefore influences the amount of packets to be sent over the network. Typical encoding standards could be e.g. H.264/MPEG-4 AVC, H.265/MPEG-H HEVC, etc. A higher encoding level corresponds to a denser compression with higher susceptibility for visible or distracting artefacts due to noise from the network and hereby deteriorates quality of the video streams.

[0047] According to an embodiment, the networking device is configured to receive video streams from multiple senders, wherein the networking device is configured to perform the buffer management per sender, per group of one or more senders, or, for one sender of focus.

[0048] The network paths between the network device and the multiple senders will likely have different characteristics in bandwidth, latency, jitter, etc. In order to accommodate to the senders, the network device possibly is configured to take into account different sender requirements and to provide different or several media encodings. Each of the senders can e.g. have corresponding low and high priority video streams, which can be visualized at the receiver side, either individually (e.g. PiP or PaP together with sender of focus) or simultaneously (e.g. grid like manner).

[0049] In case of buffer management per group of one or more senders, adapting the quality for a group of senders is advantageous as it allows for efficient multiparty conferencing or distribution. As there can be a significant amount of receivers and e.g. corresponding RTCP reports, or reports from any other supporting transport protocols, with each individual settings, grouping senders allows for a reduced set of low priority and high priority encodings to be applied and to be sent to the appropriate receiver(s). Hence, common quality for audio and/or video streams can be set for all senders in a group.

[0050] In case of multiparty conferencing, the receiver needs to deal with multiple senders in the conference. Typically, one of the senders will have the focus, whereby the other senders are of less importance in the overall conference. The notion of focus can relate to the audio stream of a sender. Nevertheless, audio streams of all contributing parties could be mixed.

[0051] In case of buffer management for a sender of focus, the networking device is configured to receive the audio stream, low priority and high priority video streams from the sender of focus and to provide feedback for these streams. The other senders can be handled in a more conservative way. The advantage is a more limited amount of control to be taken care of at receiver and sender side.

[0052] In case of buffer management per sender, the networking device is configured to receive audio streams, low priority and high priority video streams from all senders in the multiparty video conference. Possibly the network device reports back to all senders in order to optimize the communication with all senders. The advantage of this approach is that changing focus in the conference does not exhibit any effect on QoE for the receiver since all communication with all senders can be adapted optimally. Nevertheless, the amount of control increases at receiver and sender side.

[0053] In a second aspect, embodiments of the present invention relate to a method for buffer management in video streaming over a network. The method comprises the receiving of one or more video streams from a network. The one or more video streams comprise one or more low priority video streams containing low priority content and one or more high priority video streams containing high priority content. Further, the method buffers packets of the one or more low priority video streams respectively into one or more low priority buffers and buffers packets of the one or more high priority video streams respectively into one or more high priority buffers. Buffer delays of the buffers are monitored in order to determine a common delay b* for the buffers, wherein b* is smaller than a monitored low priority buffer delay if the monitored low priority buffer delay exceeds the monitored high priority buffer delays. The determined common delay b* is then enforced to the buffers. The packets of the one or more low priority and the one or more high priority buffers are combined into frames of a combined video stream.

[0054] According to an embodiment, the method further comprises the buffering of packets of an audio stream received from the network. Also the buffer delay of the audio buffer agent is monitored and the determined common delay b* is enforced to the audio buffer.

[0055] In a third aspect, the disclosure relates to a computer program product comprising computer-executable instructions for performing the method according to the second aspect when the program is run on a computer.

[0056] In a fourth aspect, the disclosure relates to a computer readable storage medium comprising the computer program product according to the third aspect.

[0057] In a fifth aspect, the disclosure relates to a data processing system programmed for carrying out the method according to the second aspect.

Brief Description of the Drawings

[0058]

Fig. 1 illustrates a networking device according to an embodiment of the invention; and

Fig. 2 illustrates steps performed by a buffer synchronization agent for the networking device in Fig. 1.

Detailed Description of Embodiment(s)

[0059] Fig. 1 illustrates a networking device 100 according to an embodiment of the invention.

[0060] The networking device 100 comprises a network interface 250 for receiving an audio stream, a region-of-interest (ROI) stream and a background (BG) stream from a network 200 via respectively separate (e.g. RTP) ports 300, 400 and 500. Hereby, differentiated services (e.g. "DiffServ" for RTP-communication) can be employed to the audio, ROI and BG stream, whereby the audio stream is received via a network path with an equal or lower congestion level than with the network path for the ROI stream, and also whereby the ROI stream is received via a network path with an equal or lower congestion level than with the network path for the BG stream. The BG video stream corresponds to low priority content and the ROI video stream corresponds to high priority content.

[0061] The networking device 100 comprises depacketizers 301, 401 and 501 for respectively depacketing the respectively audio stream, ROI stream and BG stream and subsequently provide the depacketed streams to the respective audio buffer 302, ROI video buffer 402 and BG video buffer 502.

[0062] The receiving device 100 comprises the buffer management agent 600 suitable for determining the common delay b* according to the importance of the streams when reconstructing the audio/video content at the receiver 100. The received audio stream is purposed to have the minimum buffer delay of the monitored buffer delays. In this embodiment, the ROI and BG streams comprise the necessary metadata to compose the final video output at the receiver side. These operations are performed by a video composer that takes the available decoded BG and ROI streams as input and produces frames to be displayed according to the playing device capabilities and according to the metadata on how to merge the ROI and BG decoded video stream as frames to be displayed. The resulting audio stream and video stream are played out by respectively an audio player 303 and a video player 403.

[0063] The synchronization agent 600 maintains a tuneable parameter α corresponding to a level of interactivity of the audio/video communication. The value of α is in the range of $[1, b_{max}/b_a]$ so that the negotiated common buffer delay b* never exceeds b_{max}.

[0064] The value of do corresponds to a tolerance of the end-user to the level of distortion (e.g. measured by PSNR, SSIM or MOS score evaluation). Parameter do concerns a configurable parameter, whereby a user can set or change the distortion level that he can tolerant e.g. via a user-interface.

[0065] The flowchart of Fig. 2 illustrates steps performed by the buffer synchronization agent in Fig. 1.

[0066] The three delay buffers 302, 402, 502 announce (in step 700) their respectively monitored buffer delay b_a, b_{ROI} and b_{BG} to the synchronization agent 600.

[0067] The synchronization agent 600 calculates (in step 701) a common buffer delay as $b^* = \alpha \cdot b_a$. The default value of α can be randomly chosen from the range of $[1, b_{max}/b_a]$. Its value is constantly updated in accordance to the ROI and BG buffer delay adapted in the previous round.

[0068] If $b^* \geq B_{ROI}$ (checked in step 702), i.e. the estimated common buffer delay is larger than the announced ROI buffer delay b_{ROI}, the synchronization agent 600 chooses the smaller buffer delay $b^* = b_{ROI}$ (in step 703).

[0069] If $b^* < B_{ROI}$ (checked in step 702), i.e. the estimated common buffer delay is smaller than the announced ROI buffer delay B_{ROI}, the synchronization agent first checks (in step 704) whether $B_{ROI} \geq b_{max}$. This situation typically happens with improving network condition for delivering the ROI stream. If so, b* shall be assigned to $b^* = b_{max}$ (in step 705) regardless of the distortion caused by dropping packets. This corresponds to the maximum determined buffer delay without significantly impairing the interactivity of the communication. If $B_{ROI} < b_{max}$, the common delay is $b^* = B_{ROI}$ (in step 706), so as to maintain or guarantee that the quality of the ROI stream.

[0070] In the following step, the relationship between b* and b_{BG} is examined in step 707. If $b^* \geq b_{BG}$, b* keeps unchanged (in step 708), as the quality of the ROI stream is assumed to be important to the end-user. Otherwise if $b^* < b_{BG}$, the announced BG buffer delay b_{BG} needs to be reduced. The synchronization agent 600 calculates the amount of BG buffer delay b'(d), to be reduced as a function of the imposed distortion (e.g. after error concealment or without concealment). This reduction is so that the incurred video distortion does not exceed the distortion threshold, i.e. $d \leq d_0$. The common buffer delay is thereby assigned as $b^* = b_{BG} - b'(d)$, with $b^* \geq B_{ROI}$ (in step 709).

[0071] Subsequently, the three buffers are enforced with the common buffer delay b* in step 710, whereafter α is updated in step 711.

[0072] In an alternative embodiment only b_a and b_{ROI} are considered to determine the common buffer delay b*, whereby b_{BG} is proportional to B_{ROI}. This corresponds to the flowchart of Fig. 2 without steps 707, 708 and 709.

[0073] Although the present invention has been illustrated by reference to specific embodiments, it will be apparent to those skilled in the art that the invention is not limited to the details of the foregoing illustrative embodiments, and that the present invention may be embodied with various changes and modifications without departing from the scope thereof. The present embodiments are therefore to be considered in all respects as

illustrative and not restrictive, the scope of the invention being indicated by the appended claims rather than by the foregoing description, and all changes which come within the meaning and range of equivalency of the claims are therefore intended to be embraced therein. In other words, it is contemplated to cover any and all modifications, variations or equivalents that fall within the scope of the basic underlying principles and whose essential attributes are claimed in this patent application. It will furthermore be understood by the reader of this patent application that the words "comprising" or "comprise" do not exclude other elements or steps, that the words "a" or "an" do not exclude a plurality, and that a single element, such as a computer system, a processor, or another integrated unit may fulfil the functions of several means recited in the claims. Any reference signs in the claims shall not be construed as limiting the respective claims concerned. It is to be understood that the terms so used are interchangeable under appropriate circumstances and embodiments of the invention are capable of operating according to the present invention in other sequences, or in orientations different from the one(s) described or illustrated above.

Claims

1. A networking device (100) for buffer management in video streaming over a network, said networking device (100) comprising:

- a network interface (250) for receiving one or more video streams from a network (200), said one or more video streams comprising:
  - ◦ one or more low priority video streams containing low priority content;
  - ◦ one or more high priority video streams containing high priority content;
- one or more low priority buffers (502) for buffering packets of respectively said one or more low priority video streams;
- one or more high priority buffers (402) for buffering packets of respectively said one or more high priority video streams;
- a buffer synchronization agent (600) configured to monitor buffer delays of said buffers and configured to determine a common delay b* for said buffers (402, 502), wherein b* is smaller than a monitored low priority buffer delay if said monitored low priority buffer delay exceeds the monitored high priority buffer delays, and wherein said buffer synchronization agent (600) is configured to enforce said determined common delay b* to said buffers (402, 502);
- a composer configured to combine said packets of said video buffers (402, 502) into frames of a combined video stream.

2. A networking device (100) according to claim 1, wherein said networking device further comprises:

- an audio buffer (302) for buffering packets of an audio stream, received from said network by said network interface (250), wherein said buffer synchronization agent (600) is configured to monitor the buffer delay of said audio buffer (302), and wherein said buffer synchronization agent (600) is configured to enforce said determined common delay b* to said audio buffer (302).

3. A networking device (100) according to claim 2, wherein said determined common delay b* also equals or exceeds said monitored buffer delay of said audio buffer, equals or exceeds said monitored buffer delays of said one or more high priority buffers and, equals or exceeds a minimum value b_{min}.

4. A networking device (100) according to claim 1, wherein said common delay b* maximally is determined to a value b_{max}, whereby b_{max} preferably is situated between 20ms and 150ms.

5. A networking device (100) according to claim 1, wherein if a monitored low priority buffer delay exceeds the monitored high priority buffer delays, said buffer synchronization agent (600) is adapted to determine said common delay b* sufficiently low while not exceeding a level of distortion do for said buffered low priority video stream and/or for said combined stream.

6. A networking device (100) according to claim 1, wherein said networking device (100) is configured to apply error concealment to said low priority content or frames of said combined video stream, if said determined common delay b* is smaller than a monitored low priority buffer delay and wherein said monitored low priority buffer delay exceeds said monitored high priority buffer delays.

7. A networking device (100) according to claim 1, wherein said networking device (100) is configured to provide said combined video stream and/or audio stream to a device suitable for playing said combined video stream and/or audio stream.

8. A networking device (100) according to claim 1, wherein said network interface is configured to receive one or more streams of said one or more high priority video streams over said network via one or more ports of said network interface, separate from a port for receiving one or more streams of said one or more low priority video streams, and configured

to receive said one or more streams of said one or more high priority video streams via one or more corresponding network paths with a lower congestion level than the network paths corresponding to said one or more low priority video streams.

9. A networking device (100) according to claim 2, wherein said network interface is configured to receive said audio stream over said network via a port of said network interface, separate from one or more ports for receiving said video streams, and configured to receive said audio stream via a corresponding network path with a lower congestion level than the network paths corresponding to said video streams.

10. A networking device (100) according to claim 1, wherein said networking device (100) is configured to receive a high priority video stream with a higher quality than said one or more low priority video streams, in case of degraded network congestion, wherein said quality corresponds to an encoding level, whereby said quality is inversely proportional to a congestion level of a network path of said network (200) for said stream.

11. A networking device (100) according to claim 1, wherein said networking device (100) is configured to receive video streams from multiple senders, wherein said networking device (100) is configured to perform said buffer management per sender, per group of one or more senders, or, for one sender of focus.

12. A method for buffer management in video streaming over a network (200) comprising:

- receiving one or more video streams from a network (200), said one or more video streams comprising:
    - one or more low priority video streams containing low priority content;
    - one or more high priority video streams containing high priority content;
- buffering packets of said one or more low priority video streams respectively into one or more low priority buffers (502);
- buffering packets of said one or more high priority video streams respectively into one or more high priority buffers (402);
- monitoring buffer delays of said buffers (402, 502);
- determining a common delay b* for said buffers (402, 502), wherein b* is smaller than a monitored low priority buffer delay if said monitored low priority buffer delay exceeds the monitored high priority buffer delays;
- enforcing said determined common delay b* to said buffers (402, 502);
- combining said packets of said one or more low priority and said one or more high priority buffers into frames of a combined video stream.

13. A computer program product comprising computer-executable instructions for performing the method according to claim 12 when the program is run on a computer.

14. A computer readable storage medium comprising the computer program product according to claim 13.

15. A data processing system programmed for carrying out the method according to claim 12.

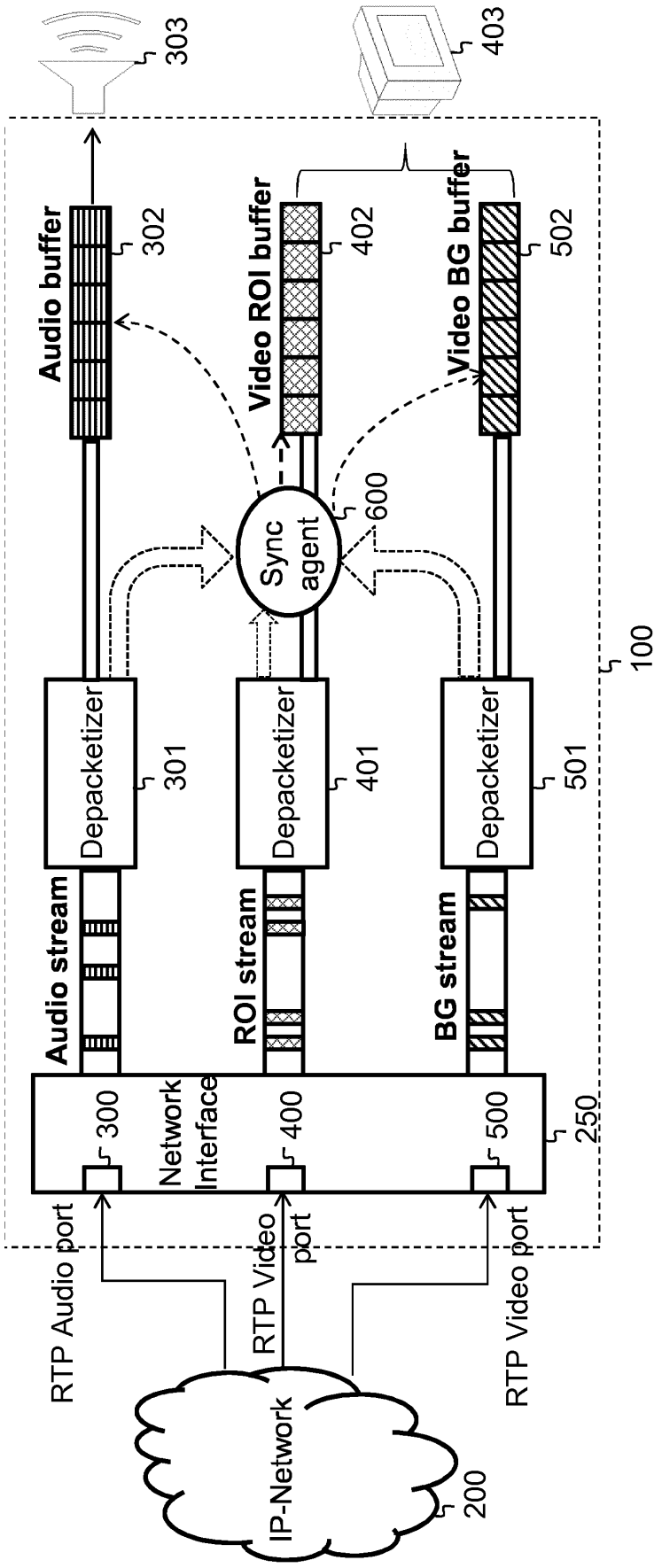

303
403
Audio buffer
302
Video ROI buffer
402
Video BG buffer
502
Sync agent
600
100
Depacketizer
301
Depacketizer
401
Depacketizer
501
Audio stream
ROI stream
BG stream
300
Network Interface
400
500
250
RTP Audio port
RTP Video port
RTP Video port
IP-Network
200

Fig. 1

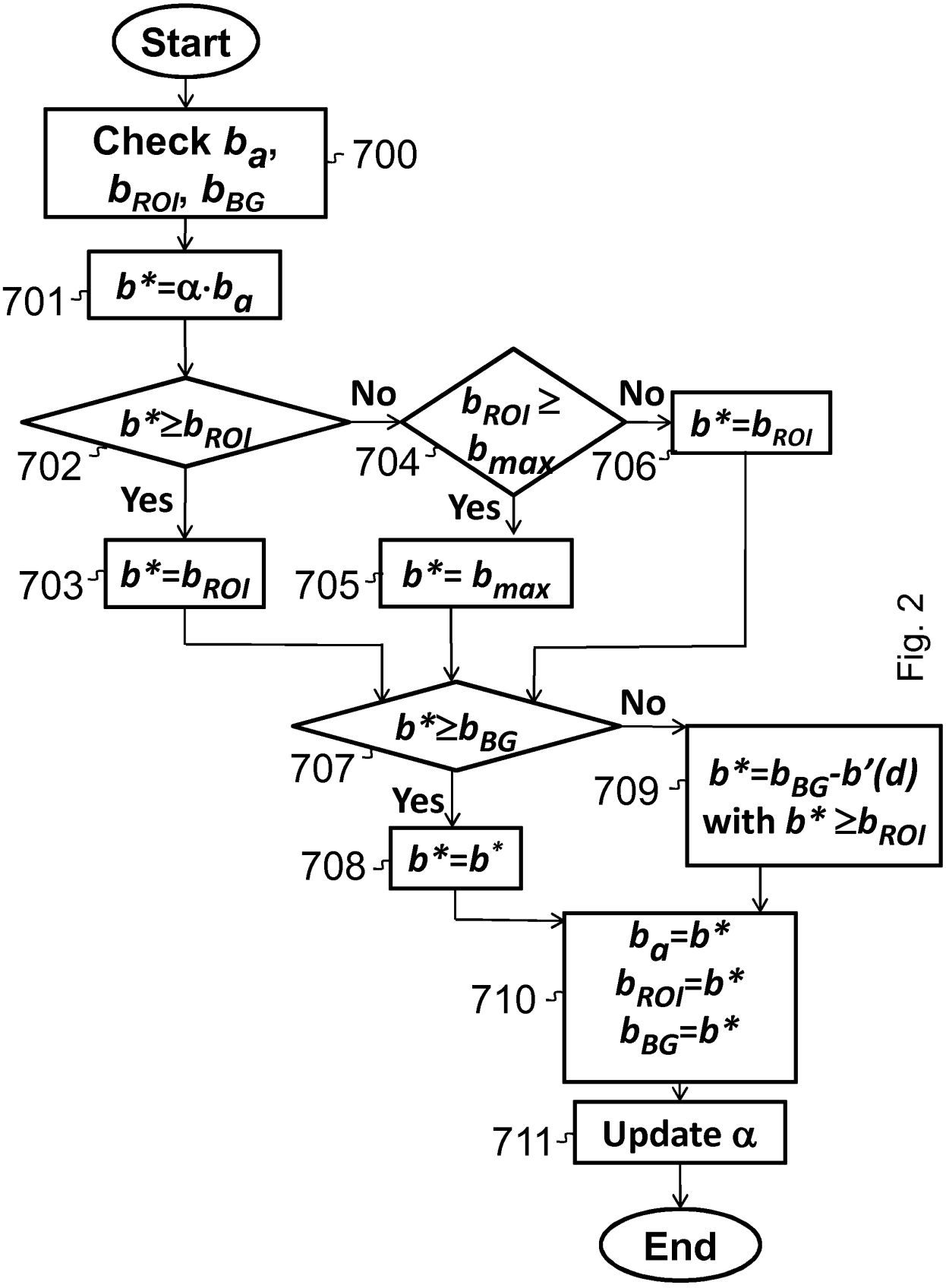

Start
Check b_a, b_{ROI}, b_{BG}
700
701
$b^*=\alpha \cdot b_a$
$b^* \geq b_{ROI}$
702
No
Yes
$b_{ROI} \geq b_{max}$
704
No
Yes
$b^*=b_{ROI}$
706
703
$b^*=b_{ROI}$
705
$b^*= b_{max}$
$b^* \geq b_{BG}$
707
No
Yes
709
$b^*=b_{BG}-b'(d)$ with $b^* \geq b_{ROI}$
708
$b^*=b^*$
710
$b_a=b^*$
$b_{ROI}=b^*$
$b_{BG}=b^*$
711
Update α
End

Fig. 2

Europäisches Patentamt
European Patent Office
Office européen des brevets

EUROPEAN SEARCH REPORT

Application Number
EP 14 29 0318

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (IPC)
A	EP 1 087 579 A2 (NTT DOCOMO INC [JP]) 28 March 2001 (2001-03-28) * page 4, paragraph 24 - page 10, paragraph 86; claim 1; figures 1, 10 *	1-15	INV. H04L29/06 H04N21/24 H04N21/439 H04N21/44
A	EP 1 511 326 A1 (SAMSUNG ELECTRONICS CO LTD [KR]) 2 March 2005 (2005-03-02) * page 4, paragraph 24 - page 6, paragraph 46; claim 1; figures 6, 8 *	1-15	H04L12/861

TECHNICAL FIELDS SEARCHED (IPC)

H04L
H04N

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
The Hague	10 April 2015	Karavassilis, Nick

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons

& : member of the same patent family, corresponding document

EPO FORM 1503 03.82 (P04C01)

ANNEX TO THE EUROPEAN SEARCH REPORT ON EUROPEAN PATENT APPLICATION NO. EP 14 29 0318

This annex lists the patent family members relating to the patent documents cited in the above-mentioned European search report. The members are as contained in the European Patent Office EDP file on
The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

10-04-2015

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 1087579 A2	28-03-2001	DE 60016347 D1	05-01-2005
		DE 60016347 T2	01-12-2005
		EP 1087579 A2	28-03-2001
		US 6922731 B1	26-07-2005
EP 1511326 A1	02-03-2005	CN 1592392 A	09-03-2005
		EP 1511326 A1	02-03-2005
		KR 20050021812 A	07-03-2005
		US 2005047417 A1	03-03-2005

EPO FORM P0459

For more details about this annex : see Official Journal of the European Patent Office, No. 12/82